Glycosyl hydrolase genes and their use for producing enzymes for the biodegradation of carrageenans

INSERT A BACKGROUND OF THE INVENTION

The present invention relates to glycosyl hydrolase genes for the biotechnological production of oligosaccharides, especially sulfated oligo-carrageenans and more particularly oligo-iota-carrageenans and oligo-kappa-carrageenans, by the biodegradation of carrageenans.

The sulfated galactans of Rhodophyceae, such as agars and carrageenans, represent the major polysaccharides of Rhodophyceae and are very widely used as gelling agents or thickeners in various branches of activity, especially agri-foodstuffs. About 6000 tonnes of agars and 22,000 tonnes of carrageenans are extracted annually from red seaweeds for this purpose. Agars are commercially produced by red seaweeds of the genera *Gelidium* and *Gracilaria*. Carrageenans, on the other hand, are widely extracted from the genera *Chondrus, Gigartina* and *Eucheuma*.

Carrageenans consist of repeat D-galactose units alternately bonded by β 1→4 and α 1→3 linkages. Depending on the number and position of sulfate ester groups on the repeat disaccharide of the molecule, carrageenans are thus divided into several different types, namely: kappa-carrageenans, which possess one sulfate ester group, iota-carrageenans, which possess two sulfate ester groups, and lambda-carrageenans, which possess three sulfate ester groups.

The physicochemical properties and the uses of these polysaccharides as gelling agents are based on their capacity to undergo ball-helix conformational transitions as a function of the thermal and ionic environment [Kloareg et al., Oceanography and Marine Biology - An annual review 26 : 259-315 (1988)].

Furthermore, carrageenans are structural analogs of the sulfated polysaccharides of the animal extracellular matrix (heparin, chondroitin, keratan, dermatan) and they exhibit biological activities which are related to certain functions of these glycosaminoglycans.

In particular, carrageenans are known:

(i) - for their action on the immune system, causing the secretion of interleukin or prostaglandins,

(ii) - for their antiviral action on the AIDS virus HIV1, the herpes virus HSV1 and the hepatitis A virus,

(iii) - as antagonists of the fixation of the growth factors of human cells,

(iv) - and also for their action on the proliferation of keratinocytes and their action on the contractility of fibroblasts.

Furthermore, oligocarrageenans act on the adherence, the division and the protein synthesis of human cell cultures, doubtless as structural analogs of the glycosylated part of the proteins of the extracellular matrix. In plants, oligocarrageenans very significantly elicit enzymatic activities which are markers of growth (amylase) or of the phenolic defense metabolism (laminarinase, phenylalanineammonium lyase).

Carrageenans are extracted from red seaweeds by conventional processes such as hot aqueous extraction, and oligocarrageenans are obtained from carrageenans by chemical hydrolysis or, preferably, by enzymatic hydrolysis.

The production of oligocarrageenans by enzymatic hydrolysis generally comprises the following steps:

1) production of a glycosyl hydrolase by the culture of a marine bacterium;

2) enzymatic hydrolysis of the carrageenan with the glycosyl hydrolase thus obtained; and

3) fractionation and purification of the oligocarrageenans obtained.

Microorganisms which produce enzymes capable of hydrolyzing iota- and kappa-carrageenans were isolated by Bellion et al. in 1982 [Can. J. Microbiol. **28** : 874-80 (1982)]. Some are specific for κ- or ι-carrageenan and others are capable of hydrolyzing both substrates. Another group of bacteria capable of degrading carrageenans was characterized by Sarwar et al. in 1983 [J. Gen. Appl. Microbiol. **29** : 145-55 (1983)]. These yellow-orange bacteria are assigned to the *Cytophaga* group of bacteria and some of these bacteria have the property of hydrolyzing both agar and carrageenans.

Purification and characterisation of several ι-carrageenases and κ-carrageenases, such as the ι-carrageenase and κ-carrageenase of *Cytophaga drobachiensis*, the ι-carrageenase of *Alteromonas fortis* and the κ-carrageenase of *Alteromonas carrageenovora*, were described in the thesis of P. Potin ["Recherche, production, purification et caractérisation de galactane-hydrolases pour la préparation des parois d'algues rouges", (February 1992)]. A detailed study of the κ-carrageenase of *Alteromonas carrageenovora* was described by Potin et al. [Eur. J. Biochem. 228, 971-975 (1995)].

The availability of specific enzymes and tools for obtaining oligocarrageenans by genetic engineering could markedly improve their production.

SUMMARY OF THE INVENTION

The Applicant has now found novel glycosyl hydrolase genes which make it possible specifically to obtain either oligo-iota-carrageenans or oligo-kappa-carrageenans.

Thus the present invention relates to novel genes which code for glycosyl hydrolases having an HCA score with the iota-carrageenase of *Alteromonas fortis* which is greater than or equal to 65%, preferably greater than or equal to 70% and advantageously greater than or equal to 75% over the domain extending between amino acids 164 and 311 of the sequence [SEQ ID No. 2] of the iota-carrageenase of *Alteromonas fortis*.

The present invention relates more particularly to the nucleic acid sequence [SED ID No. 1] which codes for an iota-carrageenase as defined above, the amino acid sequence of which is the sequence [SEQ ID No. 2].

The present invention further relates to the genes which code for glycosyl hydrolases having an HCA score with the kappa-carrageenase of *Alteromonas carrageenovora* which is greater than or equal to 75%, preferably greater than 80% and advantageously greater than 85% over the domain extending between amino acids 117 and 262 of the sequence ~~[SEQ ID No. 6]~~ of the kappa-carrageenase of *Alteromonas carrageenovora*.

In particular, the invention relates to the nucleic acid sequence [SEQ ID No. 7] which codes for a kappa-carrageenase having a score as defined above, the amino acid sequence of which is the sequence [SEQ ID No. 8].

The glycosyl hydrolase genes of the invention are obtained by a process which consists in selecting proteins having an HCA score with the iota-carrageenase of *Alteromonas fortis* which is greater than or equal to 65%, preferably greater than or equal to 70% and advantageously greater than or equal to 75% over the domain extending between amino acids 164 and 311 of the sequence [SEQ ID No. 2] of the iota-carrageenase of *Alteromonas fortis*, and in sequencing the resulting genes by the conventional techniques well known to those skilled in the art.

The glycosyl hydrolase genes of the invention can also be obtained by a process which consists in selecting proteins having an HCA score with the kappa-carrageenase of *Alteromonas carrageenovora* which is greater than or equal to 75%, preferably greater than 80% and advantageously greater than 85% over the domain extending between amino acids 117 and 262 of the sequence [SEQ ID No. 6] of the kappa-carrageenase of *Alteromonas carrageenovora*, and in

sequencing the resulting genes by the conventional techniques well known to those skilled in the art.

Finally, the present invention relates to the use of the above glycosyl hydrolase genes for obtaining, by genetic engineering, glycosyl hydrolases which are useful for the biotechnological production of oligocarrageenans.

The glycosyl hydrolases according to the invention are therefore characterized by the HCA score which they possess with a particular domain of the amino acid sequence of the iota-carrageenase of *Alteromonas fortis* or the kappa-carrageenase of *Alteromonas carrageenovora*.

The HCA or "Hydrophobic Cluster Analysis" method is a method of analyzing the sequences of proteins represented as a two-dimensional structure, which has been described by Gaboriaud et al. [FEBS Letters 224, 149-155 (1987)].

It is known that the three-dimensional structure of a protein governs its biological properties, the production of an active protein demanding correct folding.

It is also known that the primary structure of proteins varies much more substantially than the higher-order structures and that proteins can be grouped into families which show similar secondary and tertiary structures but sometimes have such divergent primary sequences that the mutual relationship between such proteins is not obvious. The code which relates primary structure and secondary structure therefore appears to be highly degenerate since very different primary structures can ultimately lead to similar secondary and tertiary structures [Structure 3, 853-859 (1995) and Proc. Natl. Acad. Sci. USA 92 (1995)].

The use of the HCA method has shown that the distribution, size and shape of these hydrophobic clusters along the amino acid sequences are representative of the 3D folding of the proteins studied.

Also, Woodcock et al. [Protein Eng. 5, 629-635 (1992)] have shown that the hydrophobic clusters defined by the α-helical 2D diagram are statistically centered on the regular secondary structures (α-helices, β-strands), that the 2D diagram based on the α-helix carries the greatest amount of structural information and that the correspondence between hydrophobic clusters and elements of secondary structure is of the same quality for any type of folding (all α, all β, α/β and α + β), thus demonstrating that the HCA method can be used irrespective of the type of protein.

L. Lemesle-Varloot et al. [Biochimie 72, 555-574 (1990)] have shown that when two proteins have a similar distribution of hydrophobic clusters over a domain of at least 50 residues, their three-dimensional structures in this domain are considered to be superimposable and their functions to be analogous.

Thus, for example, Barbeyron et al. [Gene 139, 105-109 (1994)] used this HCA method for the comparison of the similarities in the shape, distribution and size of several hydrophobic clusters of the κ-carrageenase of *Alteromonas carrageenovora* with respect to enzymes from family 16 of glycosyl hydrolases.

The two-dimensional representation used in the HCA method is an α-helix in which the amino acids are arranged by computer processing to give 3.6 residues per turn. To obtain an easily readable plane image, the helix is cut in the longitudinal direction. Finally, to obtain the whole of the hydrophobic clusters situated at the edges of the image, the diagram is duplicated. The method uses a code which recognizes only two states: the hydrophobic state and the hydrophilic state.

The amino acids recognized as being hydrophobic are identified and grouped into characteristic geometric figures. Using these two states makes it possible to become independent of the tolerance shown by the two- and three-dimensional structures towards the variability of the primary sequences. Furthermore, this representation affords rapid observation of interactions over a short or medium distance since the first amino acid and the second, adjacent amino acid of a given residue are located on a segment of 17 amino acids. Finally, in contrast to the analytical methods based on the primary or secondary structures of proteins, no "window" of predefined length is used.

The fundamental characteristic of the α-helix representation is that, for a given globular protein or only a domain of this protein, the distribution of the hydrophobic residues on the diagram is not random. The hydrophobic residues (VILFWMY) form clusters of varying geometry and size. On the diagram, the hydrophilic and hydrophobic faces of the amphiphilic helices are very recognizable. Thus a horizontal diamond cluster corresponds to the hydrophobic face of an α-helix, the internal helices appear as large horizontal hydrophobic clusters and the β-strands appear as rather short, vertical hydrophobic clusters. The method makes it possible to identify the hydrophobic residues forming the core of the globular proteins and to locate the elements of secondary structure, namely the α-helices and the β-strands, independently of any knowledge of the secondary structure of the protein studied.

The HCA score between two proteins is calculated as follows:

For each cluster:

$$\text{HCA score} = 2CR/(RC_1 + RC_2) \times 100\%$$

where

- RC_1 and RC_2 are the number of hydrophobic residues in the cluster of protein 1 (cluster 1) and the cluster of protein 2 (cluster 2), respectively.

- CR is the number of hydrophobic residues in the cluster 1 which correspond to the hydrophobic residues in the cluster 2.

The mean value obtained for all the clusters along the protein sequences compared gives the final HCA score.

On the HCA profiles, the amino acids are represented by their standard code of a single letter, with the exception of proline (P), glycine (G), serine (S) and threonine (T).

In fact, because of their particular properties, these residues are represented by the special symbols indicated below so as to facilitate their visual identification on the HCA diagrams (cf. list of abbreviations).

Proline introduces high constraints into the polypeptide chain and is considered systematically as an interruption in the clusters. In fact, proline residues stop or deform the helices and the lamellae. Glycine possesses a very substantial conformational flexibility because of the absence of a side chain in this amino acid. Serine and threonine are normally hydrophilic, but they can also be found in hydrophobic environments, such as α-helices, in which their hydroxyl group loses their hydrophilic character because of the hydrogen bond formed with the carbonyl group of the main chain. Within the hydrophobic β-lamellae, threonine is sometimes capable of replacing hydrophobic residues by virtue of the methyl group on its side chain.

Amino acids can be divided into four groups according to their hydrophobicity:

(i) - strongly hydrophobic residues: V, I, L and F;

(ii) - moderately hydrophobic residues: W, M and Y

→ W appears at surface sites more frequently than F,

→ M is encountered at various sites, internal or otherwise,

→ Y can adapt to internal hydrophobic environments and is frequently found in loops;

(iii) - weakly hydrophobic residues: A and C are virtually insensitive to the hydrophobic character of their environment; and

(iv) - hydrophilic residues: D, E, N, Q, H, K and R.

Using this HCA method, the Applicant has found that proteins having an HCA score with the iota-carrageenase of *Alteromonas fortis* which is greater than or equal to 65% over the domain extending between amino acids 164 and 311 of said iota-carrageenase are enzymes of the glycosyl hydrolase type and more particularly iota-carrageenases appropriate for the production of oligo-iota-carrageenans from carrageenans.

The proteins having an HCA score which is greater than or equal to 70%, preferably greater than or equal to 75%, with the above domain 164-311 are particularly preferred for the purposes of the invention.

One particular example of glycosyl hydrolase obtained with a gene according to the invention is the protein having the amino acid sequence [SEQ ID No. 2], extracted from *Alteromonas fortis*.

Another particular example of glycosyl hydrolase obtained with a gene according to the invention is the protein having the amino acid sequence [SEQ ID No. 4], extracted from *Cytophaga drobachiensis*.

Likewise, the Applicant has found that proteins having an HCA score with the kappa-carrageenase of *Alteromonas carrageenovora* which is greater than or equal to 75% over the domain extending between amino acids 117 and 262 of said kappa-carrageenase are enzymes of the glycosyl hydrolase type and more particularly kappa-carrageenases appropriate for the production of oligo-kappa-carrageenans from carrageenans.

The proteins having an HCA score which is greater than or equal to 80%, preferably greater than or equal to 85%, with the above domain 117-262 are particularly preferred for the purposes of the invention.

The above proteins are advantageously extracted from marine bacteria.

One particular example of glycosyl hydrolase obtained with a gene according to the invention is the protein having the amino acid sequence [SEQ ID No. 6], extracted from *Alteromonas carrageenovora*.

Another particular example of glycosyl hydrolase obtained with a gene according to the invention is the protein having the amino acid sequence [SEQ ID No. 8], extracted from *Cytophaga drobachiensis*.

As indicated previously, the genes according to the invention, coding for glycosyl hydrolases, can be obtained by sequencing the genome of bacteria which product glycosyl hydrolases, as defined above, by the conventional methods well known to those skilled in the art.

The invention further relates to the expression vectors which carry the nucleic acid sequences according to the invention, with the means for their expression.

These expression vectors can be used to transform prokaryotic microorganisms, particularly *Escherichia coli*, or eukaryotic cells such as yeasts or fungi.

The invention will now be described in greater detail by means of the illustrative and non-limiting Examples below.

The methods used in these Examples are methods well known to those skilled in the art, which are described in detail in the work by Sambrook, Fristsch and Maniatis entitled "Molecular cloning: a laboratory manual", published in 1989 by Cold Spring Harbor Press, New York (2nd edition).

A2 BRIEF DESCRIPTION DRAWINGS

The following description will be understood more clearly with the aid of Figures 1 to 4, which respectively show the following:

INSA2 a2

~~Fig. 1: The maximum similarity alignment, according to the method of Needleman and Wunsch [J. Mol. Biol. 48, 443-453 (1970)], of the amino acid sequence of the iota-carrageenase of *Alteromonas fortis* (top part) and the iota-carrageenase of *C. drobachiensis* (bottom part).~~

Fig. 2: The HCA profiles of the amino acid sequences of the iota-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*.

INSA3 a3

~~Fig. 3: The maximum similarity alignment, according to the method of Needleman and Wunsch, 1970, J. Mol. Biol. 48, 443-453, of the amino acid sequence of the kappa-carrageenase of *Alteromonas carrageenovora* (top part) and *Cytophaga drobachiensis* (bottom part).~~

Fig. 4: The HCA profiles of the amino acid sequences of the kappa-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*.

The abbreviations or special symbols used for the amino acids in the Examples below are as follows:

Glycine: ◊
Proline: *
Threonine : □
Sérine: ⊡
Alanine: A
Valine: V
Leucine: L
Isoleucine: I
Methionine: M
Phenylalanine: F
Tryptophan: W
Cysteine: C
Asparagine: N
Glutamine: Q
Tyrosine: Y
Aspartate: D
Glutamate: E
Lysine: K
Arginine: R
Histidine: H

EXAMPLE 1

The iota-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*

SECTION 1: Cloning of the genes of the iota-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*

Cytophaga drobachiensis was isolated by the Applicant from the red seaweed *Delesseria sanguinea* [Eur. J. Biochem. 201 : 241-247 (1991)]. *Alteromonas fortis* (ATCC 43554) was obtained from the American Type Culture Collection. The strains were cultivated on a Zobell medium at 25°C.

Genome libraries of the DNAs of *C. drobachiensis* and *A. fortis* were constructed.

The strain used to construct these libraries, namely *Escherichia coli* DH5α (Rec A, *endA*1, *gyrA*96, *thi*1, *hsdR*17 [rk- mk+], *supE*44, *relA*1, *lacZ*ΔM15), was cultivated on Luria-Bertani medium (LB medium) at 37°C or on a so-called Zd medium (bactotryptone 5 g/l, yeast extract 1 g/l, NaCl 10 g/l; pH = 7.2) at 22°C, to which 2% of κ-carrageenan were added.

Ampicillin (50 µg/ml) or tetracycline (15 µg/ml) was added to the agar or non-agar culture media from stock solutions prepared in 50% ethanol (to avoid solidification at the storage temperature, -20°C), except in the case of the non-recombinant strain DH5α.

The expression vector used is plasmid pAT153 described in Nature 283 : 216 (1980). This plasmid contains two antibiotic resistance genes: a tetracycline resistance gene and a gene which codes for a β-lactamase, an enzyme of the cytoplasmic membrane which degrades ampicillin.

The total DNA of *C. drobachiensis* and the total DNA of *A. fortis* were prepared by the method described by Barbeyron et al. [J. Bacteriol. 160, 586-590 (1984)].

The genomic DNAs of *C. drobachiensis* and *A. fortis* were cleaved with the restriction endonucleases *Nde*II and *Sau*3AI respectively. In fact, in the case of *C. drobachiensis*, the restriction endonuclease *Nde*II was used preferentially because the DNA of this bacterium is methylated on the C residue of the GATC sequence.

The purified DNA fragments of 5000 to 10,000 bp were cloned at the *Bam*HI site of plasmid pAT153, which cleaves the tetracycline resistance gene.

6000 clones were obtained in each of the genome libraries.

The five positive *C. drobachiensis* clones and the two positive *A. fortis* clones, which hollowed out a hole in the ι-carrageenan after one week of culture at 22°C, are referred to respectively as pIC1 to pIC5 and pIP1 to pIP2.

1. Cloning from *C. drobachiensis*

The cloning of this gene is described in detail by T. Barbeyron in the doctoral thesis examined on 28 October 1993 at the Université Pierre et Marie Curie, Roscoff.

The plasmid DNA was isolated from the above five clones by the alkaline lysis method [Nucleic Acid Res. 7 : 1513 (1979)].

The sizes and mapping of the inserts showing an ι-carrageenase activity were determined by agarose gel electrophoresis after single and double digestion of their plasmids with various restriction enzymes.

The DNA fragments were extracted from the agarose by the glass wool method.

All the plasmids obtained contain an identical *Pvu*II fragment of 3.3 kb.

This fragment was subcloned in phagemid pbluescript KSII (Stratagene) (pICP07 and pICP16).

Likewise, the internal *Nde*I fragment and a *Hind*III fragment partially comprising the *Pvu*II fragment were subcloned to give the pICN22 and pICH42 subclones, respectively.

To locate the ι-carrageenase gene, libraries were constructed from the pICP07 and pICP16 subclones in phagemid pbluescript with the aid of the exonuclease III of *E. coli*, using the "ExoIII" kit from Pharmacia.

The subclones and the ExoIII clones obtained were plated onto Zd medium solidified with ι-carrageenan.

Only the pICP16 and pICP07 clones and the ExoIII pICP074 and pICP0712 clones (obtained by degradation with ExoIII for 4 minutes and 12 minutes, respectively, from the pICP07 clone) are ι-carrageenase-positive.

2. Cloning from *Alteromonas fortis*

The DNA of the pIP1 and pIP2 clones showed inserts of 10.45 kb and 4.125 kb respectively, having a common fragment of 3 kb. These clones showed a positive ι-carrageenase activity. Different fragments were subcloned and plated as described above. However, none of the subclones obtained proved to be ι-carrageenase-positive.

SECTION 2: Determination of the nucleotide sequences of the genes coding for the ι-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*

1. Sequence of the *Cytophaga drobachiensis* gene

Plasmid pICP0712 was used to determine the nucleotide sequence of the gene responsible for the ι-carrageenase activity of *C. drobachiensis* [SEQ ID No. 3].

This nucleotide sequence is composed of 1837 bp. Translation of the six reading frames revealed only one open frame, called *cgiA*. The potential initiation codon is situated 333 bp beyond the 5'P end of the sequence.

The protein sequence [SEQ ID No. 4] deduced from the sequence of *cgiA* is composed of 391 amino acids, corresponding to a theoretical molecular weight of 53.4 kDa. The hydropathic profile of this protein shows a hydrophobic region covering the first 24 amino acids. The presence of a positively charged amino acid (Lys) followed by a hydrophobic block and then by a polar segment of six amino acids suggests that this domain could be a signal peptide. According to the analyses performed by the method of Von Heijne [J. Mol. Biol. 184 : 99-105 (1985)], the signal peptidase would cleave between valine (Val^{24}) and threonine (Thr^{25}). The mature protein devoid of its signal peptide would have a theoretical molecular weight of 50.7 kDa. The identity of the *cgiA* gene was confirmed by determination of the amino acids at the NH_2 end of the partially purified protein. The sequence obtained matches the one deduced from the nucleotide sequence. The first amino acid is situated 14 residues from the NH_2 end generated by the signal peptidase. As the presence of the two prolines following the amino acids determined by microsequencing had slightly disturbed the order of appearance of the N-terminal residues, the sequence of an internal oligopeptide, purified by HPLC after cleavage with trypsin, was established. The sequence NH_2ATYKCOOH obtained is situated near the C-terminal end of the iotase (residues 396 to 399).

2. Sequence of the *Alteromonas fortis* gene

Plasmids pIHP15 and pIHPX17, subcloned from pIP1 and pIP2, were used to determine the nucleotide sequence of the gene responsible for the ι-carrageenase activity of *Alteromonas fortis*, SEQ ID No. 1. The 2085 bp fragment contains a single open reading frame of 1473 bp, called *cgiA*. The sequence situated upstream of the initiation codon (ATG^{211}) is not a coding sequence.

The protein sequence deduced from the sequence of the *A. fortis* ι-carrageenase gene [SEQ ID No. 2] consists of 491 amino acids, corresponding to a theoretical molecular weight of 54.802 kDa. In the present case, again, the N-terminal part of the protein exhibits a high hydrophobicity, suggesting that this domain could be a signal peptide; the hypothetical cleavage site would be situated between glycine (Gly^{26}) and alanine (Ala^{27}). The mature protein devoid of its signal peptide would have a theoretical molecular weight of 51.95 kDa, corresponding to a value similar to the molecular weight obtained with the protein purified by SDS-PAGE, namely 57 kDa.

SECTION 3: Comparison of the protein sequences of the ι-carrageenases of *Cytophaga drobachiensis* and *Alteromonas fortis*

After removal of the signal peptide from each sequence, it could be seen that the sequence of the ι-carrageenase of *C. drobachiensis* has similarities to that of the ι-carrageenase of *A. fortis*.

In fact, the two sequences of iota-carrageenase have a similarity of 43.2% over the whole of the linear sequence alignment. This similarity is particularly high (57.8%) between amino acids 164 and 311 (numbering of the iota-carrageenase of *Alteromonas fortis* (Fig. 1)).

At the same time, an HCA analysis showed that the HCA score between the two proteins is 82% over a domain of 293 amino acids and reaches 90.5% in the case of said domain 164-311 (Fig. 2).

No significant similarity to other polysaccharidases known hitherto could be demonstrated.

These two enzymes therefore constitute a novel family of glycosyl hydrolases.

EXAMPLE II:

The kappa-carrageenases of *Alteromonas carrageenovora* and *Cytophaga drobachiensis*

SECTION 1: Cloning of the kappa-carrageenase genes

Alteromonas carrageenovora ATCC 43555 was obtained from the American Type Culture Collection. The strains *A. carrageenovora* and *C. drobachiensis* were cultivated under conditions identical to those mentioned in section 1 of Example I.

Likewise, genome libraries were constructed using the strain *Escherichia coli* DH5α and plasmid vector pAT153.

1. Cloning from *Alteromonas carrageenovora*

The preparation of this gene is described in detail by T. Barbeyron in the thesis cited above (cf. Example 1) and in Gene 139, 105-109 (1994).

From the genome library of *Alteromonas carrageenova*, 4 *E. coli* clones, called K1 to K4, were capable of hydrolyzing kappa-carrageenan.

Plasmids pKA1 to pKA4 were purified from the four independent clones and mapped with the aid of the restriction endonucleases *Bam*HI, *Dra*I, *Eco*RI, *Hind*III, *Mlu*I, *Pst*I, *Pvu*II, *Sal*I, *Ssp*I, *Xba*I and *Xho*I.

The presence of a 2.2 kb *Dra*I-*Hind*III fragment was noted in each plasmid.

This common fragment, which is the whole insert of plasmid pKA3, was sequenced in its entirety from plasmid pKA3.

2. Cloning from *Cytophaga drobachiensis*

From the genome library of *C. drobachiensis*, five *E. coli* clones, called pKC1 to pKC5, were capable of hollowing out a hole in the substrate. The plasmids isolated and purified from said clones were mapped with restriction endonucleases.

Internal fragments of 1100 bp and 600 bp respectively were subcloned from pKC1 in phagemid pbluescript and were called pKCE11 and pKCN6.

Plasmids pKC1, pKCE11 and pKCN6 were used to determine the nucleotide sequence of the kappa-carrageenase gene.

SECTION 2: Determination of the sequences of the genes coding for the kappa-carrageenases of *Alteromonas carrageenovora* and *Cytophaga drobachiensis*

1. Sequence of the *Alteromonas carrageenovora* gene

The number of nucleotides in the pKA3 insert is 2180 bp. Translation in the six reading frames reveals the presence of three open frames, only one of which is complete; this one separates the other two, which are only partial. All three of them are located on the same DNA strand. The second open frame, called *cgkA*, read in the third reading frame, contains 1191 bp [SEQ ID No. 5].

The translation product of the *cgkA* gene corresponds to a protein of 397 amino acids with a theoretical molecular weight of 44,212 Da (SEQ ID No. 6). The hydropathic profile of this protein shows a highly hydrophobic domain,

extending over 25 amino acids, at the N-terminal end. This domain comprises a positively charged amino acid (Lys) followed by a segment rich in hydrophobic amino acids and then by three polar amino acids. These results suggest that a signal peptide is involved. The N-terminal sequence of the protein purified from the culture supernatant was determined, thereby confirming the identity of the gene. These results indicate that the signal peptidase cleaves the protein between residues 25 and 26, which is consistent with Von Heijne's rule (-3, -1). The mature protein therefore has a theoretical molecular weight of 41.6 kDa.

2. Sequence of the *Cytophaga drobachiensis* gene

The pKC1 insert of 4425 bp contains a single open reading frame of 1635 bp, called *cgkA* (SEQ ID No. 7).

The protein translated from the kappa-carrageenase gene is a protein comprising 545 amino acids with a molecular weight of 61.466 kDa [SEQ ID No. 8].

The hydropathic profile of this protein shows a highly hydrophobic domain at the N-terminal end, suggesting that a signal peptide is involved.

According to Von Heijne's rule (-3, -1), the cleavage site of the signal peptidase should be situated between threonine and serine in positions 35 and 36 respectively, with the codon ATG^{875} as the initiation codon.

The molecular weight of the protein, calculated after removal of the signal peptide, is 57.4 kDa, which is greater than the molecular weight determined for the purified extracellular κ-carrageenase, namely 40.0 kDa.

<u>SECTION 3</u>: Comparison of the protein sequences of the κ-carrageenases of *Alteromonas carrageenovora* and *Cytophaga drobachiensis*

The κ-carrageenase of *C. drobachiensis* has a similarity of 36.1% with the κ-carrageenase of *Alteromonas carrageenovora* over the whole of the linear sequence alignment.

This similarity is particularly high between amino acids 117 and 262 (51.8%) (numbering of the κ-carrageenase of *Alteromonas carrageenovora*) (Fig. 3).

As previously, this similarity is substantiated by HCA analysis, which shows an HCA score between the two proteins of 75.4% over said domain of 145 amino acids (Fig. 4).

HCA analysis also shows that these two proteins belong to family 16 of glycosyl hydrolases, which includes endoxyglucan transferases (XET), laminarinases, lichenases and agarases. In fact, the HCA score of the two kappa-carrageenases is 67.5% with XET, 67.6% with laminarinases, 73.7% with lichenases and 71.5% with agarases.